POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)

URL: http://dergipark.org.tr/politeknik

Use of chest x-ray images and artificial intelligence methods for early diagnosis of covid-19

Covid-19'un erken teşhisi için göğüs röntgeni görüntülerinin ve yapay zeka yöntemlerinin kullanımı

***Yazar(lar) (Author(s))**: Maral A. MUSTAFA[1], O. Ayhan ERDEM[2], Esra SÖĞÜT[3]*

ORCID[1]: 0000-0002-0601-3457

ORCID[2]: 0000-0001-7761-1078

ORCID[3]: 0000-0002-0051-2271

To cite to this article: Mustafa M. A., Erdem O. A. and Söğüt E., "Use of Chest X-Ray Images and Artificial Intelligence Methods for Early Diagnosis of Covid-19", ***Journal of Polytechnic***, 28(6): 1717-1729, (2025).

Bu makaleye şu şekilde atıfta bulunabilirsiniz: Mustafa M. A., Erdem O. A. ve Söğüt E., "Use of Chest X-Ray Images and Artificial Intelligence Methods for Early Diagnosis of Covid-19", ***Politeknik Dergisi***, 28(6): 1717-1729, (2025).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: 10.2339/politeknik.1654887

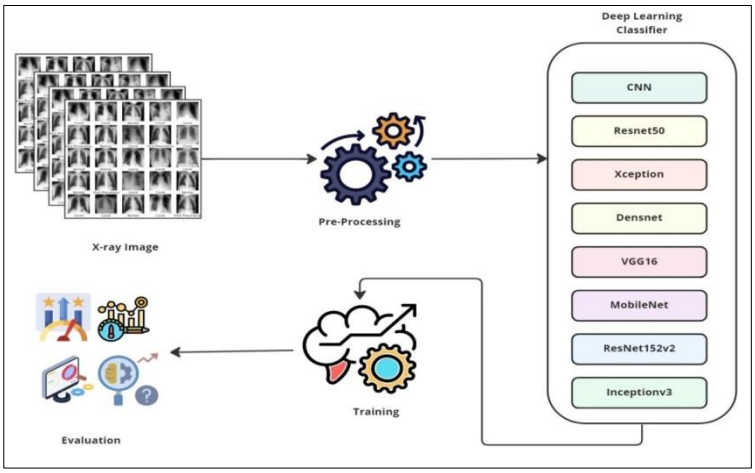
Use of Chest X-ray Images and Artificial Intelligence Methods for Early Diagnosis of COVID-19

Highlights

- *A deep learning models are proposed to classify chest X-ray images into COVID-19, Pneumonia, and Normal categories.*
- *The proposed models ensure interpretability through confusion matrices and consistent evaluation metrics.*
- *The findings are compared with existing studies in the literature to demonstrate the potential clinical relevance, architectural efficiency, and practical scalability of the proposed models.*

Graphical Abstract

A comparative deep learning frameworks are developed for the classification of chest X-ray images into COVID-19, Normal, and Pneumonia categories. The 8 CNN models are evaluated, with MobileNet and VGG16 emerging



as the most reliable and efficient solutions for clinical deployment.

Figure. Proposed deep learning-based classification approach for covid-19

Aim

To build a reliable, efficient, and interpretable deep learning framework that enables early diagnosis and differentiation of COVID-19 using chest X-ray images.

Design & Methodology

Standard preprocessing is applied to chest X-ray images and 8 deep learning models are evaluated using common performance metrics.

Originality

A multi-class, multi-model framework is presented that emphasizes standardized preprocessing, comprehensive evaluation, and clinical adaptability.

Findings

MobileNet and VGG16 achieved the highest performance, particularly for COVID-19 and Pneumonia classes.

Conclusion

This study demonstrates that pre-trained deep learning models, such as MobileNet and VGG16, can achieve high diagnostic accuracy and computational efficiency in detecting COVID-19 from chest X-ray images. These findings support their feasibility for deployment in real-time and resource-constrained healthcare environments. Future research may explore ensemble learning, integration with clinical data, and transfer learning techniques to enhance generalizability and clinical applicability.

Declaration of Ethical Standards

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Use of Chest X-ray Images and Artificial Intelligence Methods for Early Diagnosis of COVID-19

Araştırma Makalesi / Research Article

Maral A. MUSTAFA[1,2], O. Ayhan ERDEM[3], Esra SÖĞÜT[3*]

[1]Graduate School of Natural and Applied Sciences, Department of Computer Engineering, Gazi University, Ankara, Türkiye
[2]Northern Technical University, Kirkuk, Iraq
[3]Department of Computer Engineering, Faculty of Technology, Gazi University, Ankara, Türkiye

(Geliş/Received : 10.03.2025 ; Kabul/Accepted : 29.06.2025 ; Erken Görünüm/Early View :17.07.2025)

ABSTRACT

The COVID-19 pandemic has underscored the urgent need for rapid and accurate diagnostic tools to support early intervention and containment. While chest X-ray (CXR) imaging has emerged as a practical modality for COVID-19 detection, existing studies often focus on binary classification or rely on single-model evaluations without addressing class imbalance, generalizability, or multi-class diagnostic scenarios. This study proposes a novel, standardized deep learning-based framework that classifies CXR images into three clinically relevant categories: COVID-19, Normal, and Pneumonia. Unlike previous works, our approach comprehensively evaluates and compares the performance of eight prominent deep learning models—CNN, ResNet50, Xception, DenseNet, MobileNet, VGG16, ResNet152v2, and InceptionV3—using a preprocessed dataset. Key innovations include the use of a unified preprocessing pipeline, class-balancing strategy, and detailed model comparison based on a rich set of evaluation metrics (Accuracy, Precision, Recall, F1-Score, FPR, FNR, and Specificity). The results demonstrate that MobileNet and VGG16 offer high diagnostic performance with low computational overhead, making them ideal for deployment in resource-limited clinical settings. Our study's uniqueness lies in its multi-model, multi-class evaluation design, interpretability through confusion matrix analysis, and robust benchmarking against real-world challenges such as class imbalance. This positions the proposed framework as a reliable and scalable CAD tool to aid frontline healthcare providers in the early detection and differential diagnosis of COVID-19 and other respiratory illnesses.

Keywords: Covid-19 classification, vgg16, resnet, cnn, chest x-ray.

COVID-19'un Erken Teşhisi için Göğüs Röntgeni Görüntülerinin ve Yapay Zeka Yöntemlerinin Kullanımı

ÖZ

COVID-19 pandemisi, hastalıklara erken müdahale edilmesini ve yayılımın kontrol edilmesi için hızlı ve doğru tanı araçlarının gerekliliğini güçlü bir biçimde ortaya koymuştur. Göğüs röntgeni görüntüleme, COVID-19 tespiti açısından pratik bir yöntem olarak öne çıkmaktadır. Mevcut çalışmalar çoğunlukla ikili sınıflandırmaya odaklanmakta ya da tek bir modelin değerlendirmesiyle sınırlı kalmaktadır. Bu durum sınıf dengesizliği, genellenebilirlik veya çok sınıflı tanı senaryoları gibi önemli faktörleri göz ardı edilmesine sebep olmaktadır. Bu çalışma, göğüs röntgeni görüntülerini klinik açıdan anlamlı üç kategoriye ayırabilen (COVID-19, Normal ve Zatürre) yeni ve standartlaştırılmış derin öğrenme tabanlı bir çerçeve önermektedir. Önceki çalışmalardan farklı olarak, önerilen yaklaşım önişlenmiş bir veri kümesi üzerinde CNN, ResNet50, Xception, DenseNet, MobileNet, VGG16, ResNet152v2 ve InceptionV3 gibi önde gelen sekiz derin öğrenme modelinin performansını kapsamlı bir şekilde değerlendirmekte ve karşılaştırmaktadır. Çalışmanın temel yenilikleri arasında birleşik ön işleme süreci, sınıf dengesini sağlama stratejisi ve farklı değerlendirme ölçütleri kümesi (Doğruluk, Kesinlik, Duyarlılık, F1-Skoru, Yanlış Pozitif Oranı, Yanlış Negatif Oranı ve Özgüllük) durumlarına dayalı detaylı model karşılaştırması yer almaktadır. Elde edilen sonuçlar, MobileNet ve VGG16 modellerinin yüksek tanı başarımını düşük hesaplama maliyetiyle sunduğunu göstermektedir. Ayrıca bu modellerin kaynakları sınırlı klinik ortamlarda kullanım için ideal hale getirildiği vurgulanmaktadır. Çoklu model ve çoklu sınıf değerlendirme yapısı sunması, karmaşıklık matrisleri üzerinden yorumlanabilirlik sağlaması ve sınıf dengesizliği gibi gerçek dünya zorluklarına karşı bir kıyaslama sunması bu çalışmanın özgünlüğünü sağlayan ana maddelerdir. Bu özellikleriyle önerilen çerçeve, COVID-19 ve diğer solunum yolu hastalıklarının erken tespiti ve ayırıcı tanısında ön saflarda görev yapan sağlık hizmeti sağlayıcılarına destek olacak, güvenilir ve ölçeklenebilir bir bilgisayar destekli tanı aracı olarak değerlendirilmektedir.

Anahtar Kelimeler: Covid-19 sınıflandırılması, vgg16, resnet, cnn, göğüs röntgeni.

1. INTRODUCTION

COVID-19 was formally recognized as a global pandemic by the World Health Organization (WHO) on March 11, 2020, and it continues to persist due to its ongoing global transmission. Severe Acute Respiratory Syndrome (SARS), and Middle East Respiratory Syndrome (MERS) both zoonotic diseases transmitted by civets and camels, respectively, demonstrate greater

*Sorumlu Yazar (Corresponding Author)
e-mail: esrasogut@gazi.edu.tr

transmissibility yet a reduced fatality rate compared to the newly identified beta coronavirus, 2019-nCoV, which was discovered through the unbiased sequencing of patient samples [1,2]. When there are no limits between society and the natural world, new coronaviruses and diseases like SARS and MERS are more likely to appear. In the beginning, the absence of vaccination led to a rise in fatalities [3].

It was crucial to conduct study into a number of early spotting strategies to combat the COVID-19 outbreak [4]. There have been 6.49 million fatalities globally as of the WHO's most current statistics, with 30.581 in Pakistan and 104 million in the United States [5]. Today, Reverse Transcription Polymerase Chain Reaction (RT-PCR) is a technique that identifies viral RNA by analyzing samples collected via a nasopharyngeal swab. It is the gold standard for identifying coronavirus cases. The RT-PCR test's primary drawback, however, is its constrained sensitivity range, which makes it unsuitable for rapidly identifying positive instances [6].

It may be possible to prevent the spread of a coronavirus outbreak by using molecular imaging to find it early [7]. It has been proven that methods for classifying COVID-19 images mechanically often involving preprocessing, feature extraction, and selection perform more effectively than human methods [8]. As lung infections are a frequent complication of COVID-19, Computed Tomography (CT) scans can assist COVID-19 patients in locating organ harm and monitoring its development. Patients with COVID-19 may exhibit radiological imagery that is similar to those who have bacterial or viral pneumonia, such as pneumonia linked with SARS and MERS. Therefore, successfully distinguishing diseases through medical image analysis has become a difficult task that necessitates the assistance of medical staff in early disease detection and the quick isolation of infected patients [9].

A number of projects in medical image processing are concentrating on using Machine Learning (ML) methods to assist in the diagnosis of COVID-19 through the analysis of chest X-ray or CT scan images [10]. CT scans offer more comprehensive information, but they also have drawbacks, including limited accessibility, higher expense, lengthier acquisition time, and the requirement for trained radiologists [11]. The raw images are typically extracted and given to the ML technique for categorization along with manually created features like texture, shape, and point-based data [12]. Deep Learning (DL), however, has become an effective tool in the field of medical imaging [13,14]. DL needs a large quantity of data to build a model. The creation of recognition-based applications for voice recognition, object tracking, and medical imaging has been made easier by this technology [15,16].

Although additional research is needed to gain a comprehensive understanding of the epidemiological characteristics of COVID-19, the data we currently have indicates that 80% of patients have minor symptoms or are silent, while the remaining 20% have severe or critical conditions. 10% or so of patients who need to be hospitalized require admittance to an ICU for mechanical breathing. The mortality rate is said to be 2%, but some specialists believe it might actually be closer to 0.5%.

Pneumonia, an inflammation of the pulmonary parenchyma responsible for gas transmission in the lung, can be brought on by COVID-19. Pneumonia is not a singular illness but rather a collection of infections brought on by numerous viruses, bacteria, or fungus [17].

The task at hand is to use intelligent methods to categorize X-ray images into numerous classifications, including COVID-19, normal, and viral pneumonia. Advanced ML algorithms and computer vision techniques must be used for this job in order to correctly pull pertinent characteristics from the X-ray images and categorize them into the appropriate categories. The identification of discriminative characteristics, handling the class imbalance, and creating an intelligible and robust categorization model maybe some of the challenges that come with the job of building such a model, considering the complexity and variability of X-ray images. Swift identification and treatment of COVID-19 and other respiratory illness could greatly benefit from the use of such a model.

The primary objective of this study is to develop an efficient and robust DL-based framework for the automated classification of chest X-ray (CXR) images into three distinct categories: COVID-19, normal, and viral pneumonia. This research aims to support early and accurate diagnosis of COVID-19 and related respiratory diseases, leveraging Convolutional Neural Networks (CNN) and different DL algorithms to enhance diagnostic precision, reduce reliance on manual interpretation, and accelerate clinical decision-making.

To summarize the main contributions of this study:

- To develop a comprehensive DL framework for the multi-class classification of chest X-ray images, aimed at distinguishing COVID-19 from other lung conditions such as pneumonia and normal cases.
- To implement and evaluate multiple pre-trained CNN architectures, fine-tuned on a curated and preprocessed dataset to enhance model performance and generalizability.
- To introduce a standardized preprocessing pipeline involving image resizing, and normalization to improve data quality and minimize model bias.
- To conduct extensive experiments assessing model performance using metrics such as Accuracy, Precision, Recall, and F1-Score, thereby offering comparative insights into the effectiveness of each architecture.

2. RELATED WORKS

For the purpose of detecting and categorizing COVID-19 using CXR images, computer vision experts have created a variety of methods in the past [18,19]. However, while the majority of studies concentrated on conventional methods, a small number of scholars created novel DL models for the detection and recognition of coronavirus from CXR and CT images [20,21]. The paradigm described by Khan et al. [22], which used deep explainable Artificial Intelligence (AI) for coronavirus categorization from CXR images, is one illustration of a DL-based strategy. For the feature capture and training phases of this system, two DL models were used. The authors enhanced feature fusing using canonical correlation analysis and optimized merged features using a mix of whale-elephant herding and feature selection. On three openly accessible datasets, the authors got accuracies of 99.1%, 98.2%, and 96.7%, which were superior to the outcomes attained by earlier techniques. The optimization method used in this study, however, had a static cutoff value, which is a drawback that might be fixed in subsequent studies.

Researchers in [23] proposed a CAD system for tuberculosis detection using chest X-rays by jointly addressing segmentation and classification tasks. They introduced TB-UNet for precise lung segmentation and TB-DenseNet for disease classification, achieving an F1-score of 0.8988 and accuracy of 0.9510, respectively.

The paradigm described by authors in [24], which used CNN-LSTM for corona categorization, is another method for CXR image analysis. Modified EfficientNetB0 was used by authors in [24] to create a new CNN-LSTM technique for deep feature extractions. The extracted features were combined using serial-based maximum value fusing methods, and the combined vector was then used to enhance moth flame feature selection. Three openly accessible databases were used in the research, which had accuracy rates of 93, 94.5, and 98.5 percent consecutively. A limitation of this study was that the fusion procedure limited the vector capacity, and lengthened computation times. For the identification of CT scan images, authors in [25] presented a new branch model network that made use of CNN and transformations. They put into practice two models with branches one with CNN and the other with transformation-based branches. A comprehensive COVID-19 dataset was utilized for the analysis, and a bi-directional approach was employed to integrate the features. The accuracy rate was 96.7%. The lack of sufficient characteristics and the insufficient patient information, however, limited this study. Using 2 chest CT images datasets, authors in [26] developed an effective CNN models for the identification of COVID-19. The highest accuracy rates were obtained with EfficientNetB5 (97.59%, and 98.45%.). They used GradCam graphics to see the weights in their multi-layer CNN model.

Researchers in [27] proposed a DL-based system to distinguish between normal and severe pneumonia cases using chest X-rays, addressing challenges such as visual similarity with other respiratory diseases and variability in image quality. They evaluated eight pre-trained models are ResNet50, ResNet152V2, DenseNet121, DenseNet201, Xception, VGG16, EfficientNet, and MobileNet on two large datasets. Among them, MobileNet achieved the highest accuracy, with 94.23% and 93.75% on the respective datasets. Hyperparameters such as batch size, epochs, and optimizers were fine-tuned to identify the most effective model configuration.

Using CXR images, authors in [28] developed a multi-classification technique for COVID-19. Features were obtained from the Global Average Pooling layer and used in a pre-trained XceptionNet model that was learned using transfer learning. The study used a three-class open source cohort with a 99.3% accuracy rate. This method's drawback was the insufficient number of images used in the chosen selection.

There are studies using DL-based techniques to diagnose COVID-19 using medical images. A hosted cuckoo optimization method was used by authors in [29] to adjust hyperparameters and arrive at a precise diagnosis. CNN with Social Mimic Optimization (CNN-SMO), SVM classifier using Bayesian Optimization algorithm (SVM-BOA), and the proposed DBN+HO-COA approach were used in the study. The accuracy rates of these models are approximately 63%, 45% and 95%. In order to increase accuracy, authors in [30] created an automatic CNN-based system for binary and multiclass categorization of chest images. To facilitate the diagnosis of COVID-19 using CT images, authors in [31] created a new CNN model called CTnet-10, which obtained an accuracy of 82.1% for a two-class dataset and 94.52% for conventional DCNN networks. For the categorization of coronavirus illness from lung ultrasound images, authors in [32] demonstrated a multi-layer fusion method using convolutional connections, obtaining a 91.8% accuracy rate but with a large number of parameters. For the categorization of CXR images, authors in [33] created an innovative CNN architecture based on 22 layers and attained high accuracies of 99.1% for binary and 94.2% for multiclass datasets.

In their research, authors in [34] improved a DL network for COVID-19 diagnosis on a binary class dataset using the Gravitational Search Algorithm method, and they reported a 98% accuracy rate. Authors in [35]'s method for COVID-19 classification using medical CXR images was built on Generative Adversarial Networks (GAN), and it achieved an amazing accuracy of 99.78%. Authors in [36] introduced a GAN-based system for the automatic segmentation and detection of COVID-19 lung infections in CT-scan images, obtaining 98.10% classification accuracy and 81.11% dice coefficient accuracy. Using transfer learning and four distinct models, authors in [37] proposed using X-ray images as a tool for COVID-19 detection. VGG16 and VGG19 performed better than the other two models, with a 99.3% success rate (accuracy)

for their approach. However, the writers did not include any other respiratory illnesses; they only utilized COVID-19 and normal images from various databases. As a result, their system was unable to identify additional lung illnesses. A technique for identifying COVID-19 infection using chest radiography images was put forth by authors in [38] using a model made up of nine Convolutional Layers (CL) and one completely connected layer. The model had an accuracy of 98.40% after being trained on multiclass datasets with three classes (COVID-19, normal, and pneumonia). A new DCNN model called COVIDXception-Net was created by authors in [39] for classifying COVID-19 illness using X-ray images. On four publicly accessible datasets, the system obtained 99.4% accuracy after the model was trained using Bayesian optimization. GRAD-CAM rendering was used for qualitative analysis.

Researchers in [40] proposed a performance analysis of ResNet50, DenseNet121, and Inception-ResNet-v2 for COVID-19 detection using chest X-rays. DenseNet121 achieved the highest internal accuracy at 96.71%, while Inception-ResNet-v2 reached 76.70% externally. To reduce performance drop across datasets, an ensemble method was introduced, improving accuracy to 97.38% (internal) and 81.18% (external).

Researchers in [41] proposed a DL framework for COVID-19 detection and severity assessment using chest X-ray images. U-Net was employed for lung segmentation, achieving a high precision of 0.9924. In the classification task, the convolutional capsule network achieved true positive rates of 86% for COVID-19, 93% for pneumonia, and 85% for normal clases. For severity prediction, models including ResNet50, VGG-16, and DenseNet201 were utilized, with DenseNet201 demonstrating the highest accuracy among them. The results, validated with 95% confidence intervals, demonstrate the framework's robustness for clinical deployment.

Researchers in [42] proposed multiple DL models to detect pneumonia from chest X-ray images, focusing on accuracy, precision, recall, loss, and AUC scores. They evaluated Enhanced CNN, VGG-19, ResNet-50, and fine-tuned ResNet-50 using an expanded dataset of 5.863 images from Kaggle. Results showed that Enhanced CNN achieved the highest accuracy at 92.4%, outperforming ResNet-50, which scored 82.8%.

Researchers in [43] proposed two novel DL algorithms - Dynamic Co-Occurrence Grey Level Matrix (DC-GLM) and Contextual Adaptation Multiscale Gabor Network (CAMSGNeT) - to improve COVID-19 detection in chest X-rays by addressing challenges such as interpretability, computational cost, and data dependency. DC-GLM captures key COVID-19 indicators like ground-glass opacities and fibrosis by modeling texture patterns and spatial pixel correlations, while CAMSGNeT enhances fine feature extraction using a Contextual Adaptive Diffusion mechanism. These methods preserve important details such as air bronchograms and improve edge and texture recognition. Combined with a lightweight neural network and feature importance analysis, the approach achieved 98.27% and 100% accuracy on two datasets, offering a highly interpretable and efficient solution.

Researchers in this paper [44] proposed an automated pneumonia detection approach using chest X-ray images by integrating features from three optimized pre-trained CNN models with an XGBoost classifier. The method combines learned features through an ensemble strategy, with Bayesian optimization used to fine-tune hyperparameters and preserve essential layers. This integration achieved strong performance, with a classification accuracy of 99.15%, precision of 99.53%, sensitivity of 99.30%, and an AUC of 0.9972.

A wide range of DL-based approaches has been proposed for the automatic detection of COVID-19 and similar respiratory diseases using CXR images. Beyond traditional methods, several studies have achieved high accuracy rates by employing specially designed CNN, CNN-LSTM, and DCNN architectures. In particular, techniques such as transfer learning, feature fusion, Bayesian optimization, and ensemble strategies have been utilized to enhance model performance, with some studies reporting accuracy and AUC values exceeding 99%. However, certain models still exhibit limitations such as restricted dataset sizes, vector capacity constraints, and lack of clinical information. In next-generation models, efforts have been made to improve both accuracy and interpretability through the use of techniques like GRAD-CAM, capsule networks, GAN-based segmentation, and even multi-modal imaging (CXR and CT).

Despite the extensive research, efforts and high classification accuracies reported in previous studies, several limitations persist. Many approaches rely heavily on large, curated datasets, making their performance less generalizable to real world, diverse clinical data. Some models suffer from limited interpretability, computational inefficiency, or lack of robustness across different imaging conditions and patient populations. Fusion and optimization techniques often increase model complexity and computation time without guaranteeing consistent gains. Moreover, a number of studies focus exclusively on binary classification (e.g., COVID-19 vs. Normal), overlooking other respiratory diseases like pneumonia, which limits their diagnostic utility in multi-class scenarios. Lastly, few works perform comprehensive comparisons across multiple DL architectures on balanced datasets, and many do not address model reproducibility or the need for lightweight deployment in clinical environments.

3. METHODOLOGY

3.1. Dataset

The dataset (CXR) displayed in Figure 1 consists of 317 X-ray images [40,45] that are separated into three categories; normal (90 images), Covid (137 images), and pneumonia (90 images).

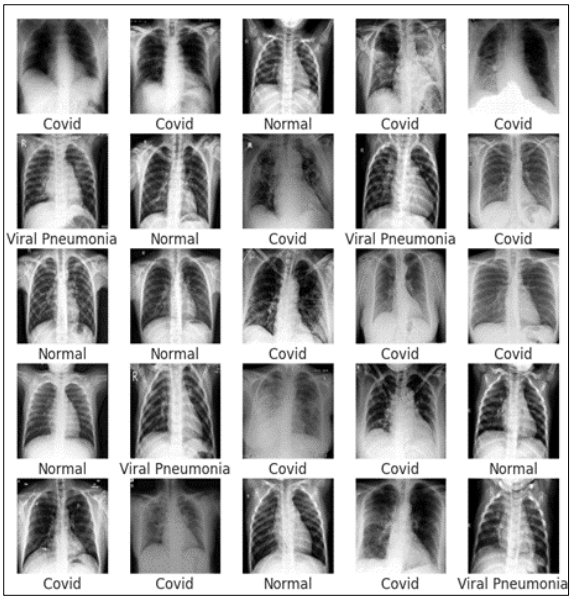


Figure 1. Some examples of images of the dataset

In this study, a series of preprocessing steps were applied to chest X-ray images used for model training and testing. The images were read from structured directories, mapped to class labels (Covid, Normal, Pneumonia), and resized to 224×224 pixels. All images were converted to RGB format and cast to the float32 data type. To prevent the model from learning any ordering bias, the training dataset was shuffled. Class distributions in both training and testing sets were computed and visualized to detect potential imbalance. Finally, pixel values were normalized from the 0–255 range to a 0–1 scale to improve training efficiency and convergence stability.

This dataset has been made available to support researchers and medical professionals in diagnosing respiratory diseases, particularly those associated with COVID-19 [40-45]. To assess the efficiency of current and future ML models in accurately classifying these images, the dataset has been partitioned into training and testing sets using an 8:2 ratio. Specifically, 80% of the data is allocated for training, enabling the model to learn patterns, while the remaining 20% is reserved for testing to evaluate the model's effectiveness on unseen data.

Chest X-ray images categorization that is accurate and trustworthy is crucial for identifying and managing respiratory illnesses, especially those brought on by Covid-19. This kind of big collection can be used to build ML models that can be used to automate diagnosis, possibly increasing both speed and accuracy. However, the caliber and variety of the training data have a significant impact on how well these algorithms work. Therefore, the advancement of the creation of ML algorithms for diagnostic reasons in healthcare depends on databases like this.

3.2. Training Details

This section presents a detailed explanation of the methodology adopted for the proposed DL-based classification approach, as depicted in Figure 2. The approach begins with the utilization of a publicly available chest X-ray dataset. In accordance with standard ML practices, the dataset is divided into two subsets: a training set and a testing set. The training set is used to build and optimize the model, allowing it to learn patterns and relevant features necessary for accurately classifying images into predefined categories (e.g., COVID-19, pneumonia, and normal). During this training phase, model parameters - including weights and biases - are iteratively adjusted to minimize prediction error.

After the training phase is completed, the model's performance and ability to generalize are assessed using the test set, which consists of previously unseen data. This step verifies that the model does not merely memorize the training data but can accurately predict outcomes on novel inputs. To quantify performance, standard evaluation metrics such as accuracy, precision, recall, and F1-score are employed.

The dataset is randomly split in to 80% of the data is used for training and 20% for testing. This split is chosen to ensure that the model is exposed to a sufficient volume of training data to learn discriminative features, while retaining enough testing data for reliable performance assessment. Additionally, class distribution is balanced to minimize bias and ensure consistent representation across all categories, thereby enhancing the fairness and robustness of the model.

Random state helps to shuffle the examples in the dataset to be flipped helping it to reduce the susceptibility bias of the dataset to the performance of the model. In this project, for reproducibility, we used a random state of 25, which helps to generate the same shuffle sequence if the code is run several times. After this, the pixel values of the images were shuffled, and normalization of the pixel values followed by dividing them by 255.0. Normalization of pixel values is a regular practice in image processing and helping to enhance the performance of DL models. It reduces the range of high numbers between 0 and 1 and helps models to easy to understand. A bar chart was then plotted to explain the number of examples in each class (Figure 3). It helped us to know if there were any discrepancies in the dataset, and if there were, we had to correct the model. Accordingly, in the test phase, the distributions of the classes are close to each other.

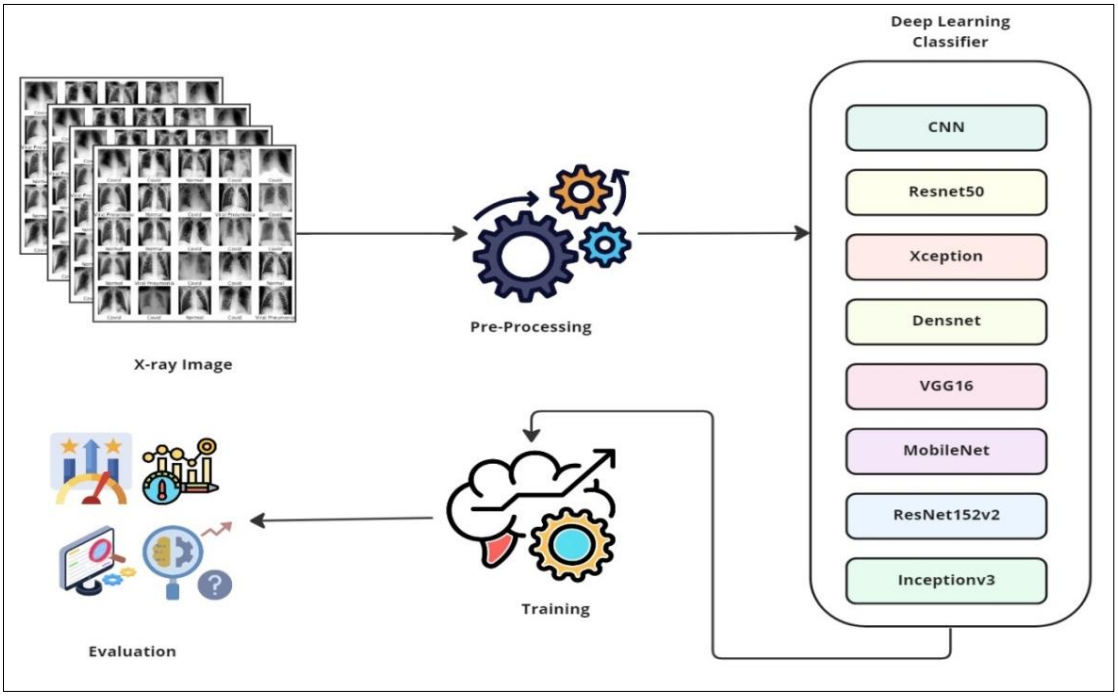


Figure 2. Our proposal of dl-based classification approach for covid-19

Covid, Normal and Pneumonia classes consist of approximately 20 images. In the train phase, the Covid class has more data than the other two classes. Normal and Pneumonia classes contain a similar number of images.

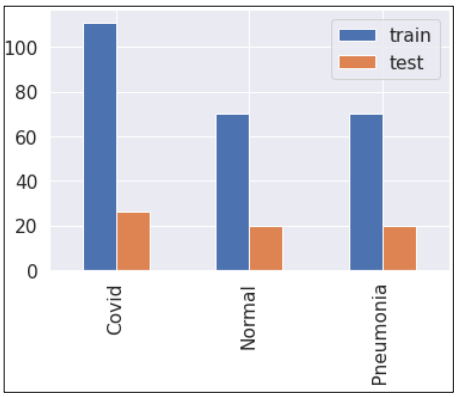


Figure 3. Number of examples in each class

A dataset with too many examples of one class and too little of another is said to be biased. Such a dataset is biased because the model can become more accurate for some classes and less accurate for others. The performance of the models could be improved by expressing each class fairly in the dataset by displaying the number of samples in the three classes; Normal, Covid, and Pneumonia. The DL techniques used to train the models are DL techniques that have frequently proven successful for computer vision problems.

3.3. Model Architecture

A well-known DL model called CNN is made to handle data with a grid-like structure, like images. The CNN design has been modified and Resnet50, Xception, DenseNet, MobileNet, VGG16, Resnet152v2, and Inceptionv3 models has been produced, each of which has been tailored for a particular job. The reason for choosing these models in the study is that their effectiveness has been proven in previous image classification studies and they are widely used in the literature. Due to the nature of this study and the limited analysis resources, the evaluations were limited to these common and proven architectures.

Using the training collection and the preprocessed data, the models were trained. By changing the weights of its synapses, the model trains itself to find trends in the data. Minimizing the discrepancy between the model's predicts, and the real labels of the training data is the objective. The models used in the study are as follows:

- CNN: The DL model has the following layers: an input, a Convolutional Layer (CL), a max pooling, a second CL, a second max pooling, a dropout, a flatten, and two thick layers. The input layer will take 224x224 images with three color channels as input (RGB). 32 3x3 filters with the ReLU Activation Function (AF) are applied in the first CL. A max pooling procedure with a pool capacity of 2x2 is carried out by the first max pooling layer. The 32 3x3 ReLU-activated filters from the second CL are applied. With a pool capacity of 2x2, the second max pooling layer executes a max pooling procedure. In order to avoid overfitting, the dropout layer arbitrarily removes 80% of the input units. The result from the prior layer is flattened into a 1-D vector by the flatten layer. 128 units make up the first thick layer, which employs the ReLU AF. Three units -

one for each class - make up the second thick layer, which employs the Softmax AF.

- ResNet50: First, three dense layers with rising amounts of nodes (2024, 2024, and 1024) and ReLU AFs are applied to the output of a pretrained model. To avoid overfitting, dropout layers are introduced after the second and third thick layers. Then, two more thick layers are added, each with 512 nodes and ReLU AFs, and then a dropout layer. The output layer also has a Softmax AF and three nodes that correlate to the dataset's three classifications. accuracy is used as the assessment measure, and the model is built using the Adam algorithm along with the sparse categorical cross-entropy Loss Function (LF). On the collection of chest X-ray images, this algorithm can be used to train the model and assess how well it performs.
- Xception: The Flatten layer is used to first level the output of the basic model. Then, a number of completely linked levels are included, each containing 2048, 1024, 512, 256, and 128 neurons. A dropout layer comes after each completely linked layer to avoid overfitting. The third and concluding layer is a dense layer with 3 neurons that outputs class odds using the Softmax AF. Since the indices are numbers, the sparse categorical cross-entropy LF is used when building the model using the Adam algorithm. During training, the model's precision is also assessed.
- DenseNet: The output of the Densenet model is then cycled through a number of completely connected layers, each of which has a dropout layer to avoid overfitting and a ReLU AF. The dropout rate falls from 0.1 to 0 as a measure of the density of neurons progressively declines from 2048 to 512. A completely connected layer with three neurons (one for each class) and a Softmax AF make up the output layer, which outputs the class odds.
- MobileNet: The input layer of the pre-trained model serves as the input layer for the new model, while its output layer is linked to a newly added dense layer with 2024 units and a ReLU activation function. To assist prevent overfitting, the next stage is to add a dropout layer at a rate of 0.1. A dropout layer with a 0.1 dropout rate is introduced after a dense layer with 2024 units and a ReLU AF. A third dense layer with 1024 units and a same AF is then added, which is followed by a further dropout layer with a rate of 0.1. A dropout layer with a rate of 0.5 is added after the introduction of a fourth thick layer with 512 units and a ReLU AF. The output layer is then given three units and a Softmax AF is utilized to generate the predicted probabilities for the three categories. The model is then built using the Adam method, sparse category Cross-Entropy LF, and precision.
- VGG16: The result of the VGG16 model is smoothed before being run through two dense layers with rates of 0.5 dropout regularization and ReLU AF. A Thick layer with a Softmax AF for multi-class categorization with three output classes makes up the final layer. The compiled model is trained with a sparse categorical cross-entropy LF and accuracy measure using the Adam algorithm. This model has an output layer with Softmax activation as the final layer, three dense layers with ReLU activation, and a total of three dense layers. Given that this is a multi-class classification problem having three classes. This first dense layer consists of 1024 neurons, 512 neurons in the second dense layer, and three neurons in the ultimate output layer.
- ResNet152v2: Four completely connected (dense) layers with ReLU AF and dropout regularization are applied to the ResNet152V2 model's output, with the first three layers' dropout rates being 0.1 and the final layer's being 0.5. The last layer is a Thick layer with multi-class categorization and a Softmax AF with three output classes. The compiled model with accuracy to sparse categorical Cross-Entropy LF was trained with the Adam algorithm. The final output layer with Softmax activation, which comprises four dense layers with ReLU activation, consists of four dense layers. The last dense layer with has 512 neurons while the first three dense layers have in total 2024 neurons.
- InceptionV3: Four completely connected layers with ReLU AF and dropout regularization are implemented on InceptionV3 model output, that functioning has rates 0.1 for the first three layers and 0.5 for the last layer. The final layer is a Thick layer with a Softmax AF designed for multi-class categorization that has three output classes.

The selection of this DL models was guided by their demonstrated effectiveness in image classification tasks, particularly within the domain of medical imaging. CNNs serve as the foundational architecture for image analysis due to their ability to extract spatial hierarchies of features. Building upon CNNs, architectures such as ResNet and DenseNet introduce residual and dense connectivity, respectively, to alleviate vanishing gradient issues and enhance feature reuse, making them highly suitable for learning complex patterns in chest X-rays. VGG16, although deeper and computationally intensive, is known for its simplicity and strong performance on smaller datasets. MobileNet is selected for its lightweight architecture optimized for performance on resource-constrained environments, which is crucial for deployment in clinical settings. InceptionV3 and Xception incorporate depthwise separable convolutions and inception modules, enabling the network to capture

multi-scale features effectively. The inclusion of these diverse models allows for a comprehensive evaluation of different architectural paradigms, ensuring that the final selection is based on empirical performance and computational efficiency within the context of COVID-19 and pneumonia detection from chest X-ray images.

The compiled model is trained with a sparse categorical cross-entropy LF, and accuracy measure using the Adam algorithm. The model has a final output layer with Softmax activation, four dense layers with ReLU activation, and four dense layers overall. The concluding dense layer has 512 neurons, while the first three Dense levels have a total of 2024 neurons.

The testing set was used to assess the models' efficiency and accuracy. A portion of the data that the algorithm has never seen before is the testing set. We can determine the model's generalizability to brand-new, untested data by assessing it on this collection of data. Accuracy, Precision, Recall, and F1-score are some of the measures used to evaluate the models' success. In addition to these, False Positive Rate (FPR), which is the rate at which samples that do not actually belong to that class are incorrectly assigned to that class, and False Negative Rate (FNR), which is the rate at which samples that actually belong to that class are incorrectly assigned to the class, are also considered as evaluation metrics. These measures are frequently employed in categorization tasks to rate the accuracy of the model's predictions.

4. EXPERIMENTAL RESULTS

The given findings (Table 1, and Figure 4) are evaluation measures for various DL models that were trained on a specific dataset. A simple CNN is the first model, and it obtained validation accuracy of 0.9216 and validation loss of 0.3166. The second model, Resnet50, which is a variation of the Residual Network design, obtained validation accuracy of 0.67 and validation loss of 0.7092. The third model, called Xception, had a loss of 0.3420 and an accuracy of 0.87. It is a deep CNN with an Inception-like design. With a loss of 0.1127 and an accuracy of 0.89, the fourth model, DenseNet, a deep neural network with tightly linked blocks, performed the best.

With a validation loss of 0.3360 and a validation accuracy of 0.924, the fifth model, MobileNet, a lightweight deep neural network made for mobile and embedded devices, performed well. A deep CNN with 16 layers and a loss of 0.0142, and a precision of 0.924 make up the sixth model, called VGG16. A more advanced version of the Residual Network design, the seventh model, Resnet152v2, obtained an accuracy of 0.87 and a loss of 0.00092. Inceptionv3, the eighth model, is a deep CNN with a design resembling that of the original Inception. It obtained a loss of 0.0268 and an accuracy of 0.83.

As a result (Table 1, and Figure 4) each model has attained different degrees of accuracy and loss, according to the assessment metrics given for the various DL models trained on the dataset. On the validation collection, some models, like MobileNet and VGG16, attained high accuracy, whereas Resnet50 and Inceptionv3 attained lesser accuracy. The various model designs, including CNN, Resnet, and DenseNet, also play a role in the variations in precision and loss that can be obtained. Ultimately, the assessment measures emphasize how crucial it is to pick the best architecture and adjust the model parameters to obtain the best possible results for a specific dataset.

The findings of the evaluation of various DL models for the classification of COVID-19 into three classes Covid (0), Normal (1), and Pneumonia (2) are displayed in the confusion matrices (Figure 5) that are given. The actual class of the instances is displayed in the appropriate row, and each cell in the matrix reflects the number of instances predicted by the model for a specific class.

All cases in the Covid and Normal classes were properly classified by the first model, CNN, while only four instances in the Pneumonia class were incorrectly classified (Figure 5). More misclassifications were observed in the second model, ResNet50, including four misclassified instances in the Normal class and eighteen in the Pneumonia class, while the Covid class was perfectly classified.

The confusion matrix for the Xception model (the third model) showed perfect classification for the Pneumonia class. In contrast, three misclassifications were observed in the Covid class (two as Normal, one as Pneumonia) and five errors in the Normal class, all being misclassified as Pneumonia.

When the confusion matrix of the DenseNet model (the fourth model) is evaluated, it is seen that high accuracy rates are achieved in all classes. Only one error was made in the Covid class. The confusion of 5 samples in the Normal class with Pneumonia showed that it is relatively more difficult to distinguish between these two classes. The Pneumonia class was also predicted with 95% accuracy.

Table 1. Loss and accuracy results according to the models

Model	Loss	Accuracy
CNN	0.3166	0.9216
Resnet50	0.7092	0.67
Xception	0.3420	0.87
DenseNet	0.1127	0.89
MobileNet	0.3360	0.924
VGG16	0.0142	0.924
Resnet152v2	0.00092	0.87
Inceptionv3	0.0268	0.83

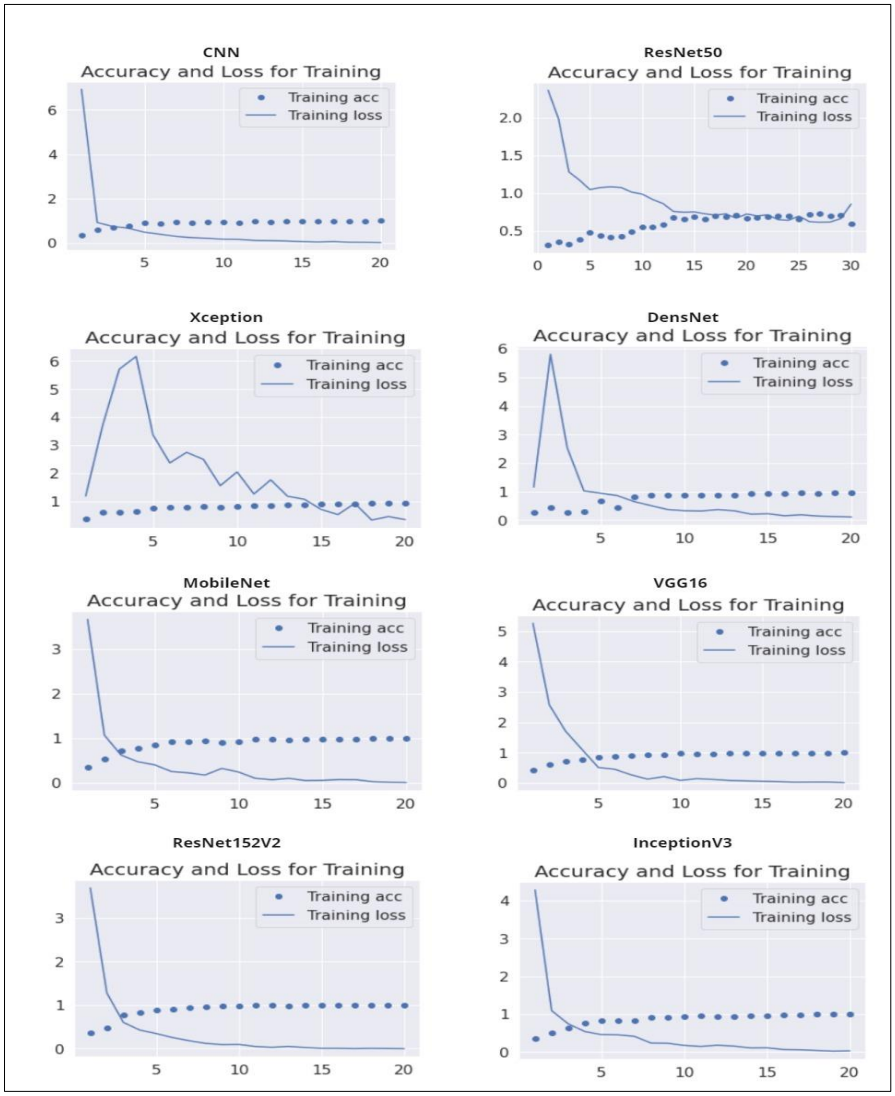


Figure 4. Accuracy and loss curves of different dl models

The confusion matrix for the MobileNet model (the fifth model) indicated perfect classification accuracy (100%) for both the Covid and Pneumonia classes, with all respective sampled correctly predicted. However, five instances from the Normal class were misclassified as Pneumonia, resulting in a slightly lower accuracy (75%) for that class.

The confusion matrix for the VGG16 model (the sixth model) showed perfect classification for both the Covid and Pneumonia classes, achieving 100% accuracy in both. However, five instances from the Normal class were misclassified as Pneumonia, resulting in a 75% accuracy for the Normal class.

The confusion matrix for the ResNet152v2 model (the seventh model) indicated high but not perfect classification performance across all three classes. The model achieved an accuracy of 92.3% for Covid, 80% for Normal, and 90% for Pneumonia.

The confusion matrix for the Inceptionv3 model (the last model) revealed notable misclassifications across all three classes. While the model achieved 92.3% accuracy for the Covid class, the performance was lower for Normal (75%) and Pneumonia (80%) classes.

Generally, the analysis of the confusion matrices reveals that some models are better adapted to the three-class classification task (i.e., COVID-19, Normal, and Pneumonia), achieving high accuracy with minimal misclassifications. In contrast, other models exhibit significant confusion between classes and may benefit from further optimization or alternative architectural choices.

The findings (Table 2) demonstrate how various DL models perform when given the job of categorizing chest X-ray images into three groups: Covid, Normal, and Pneumonia. Precision, recall, and F1-score are the metrics that are used to assess the models and give information on how well they work generally and for each class.

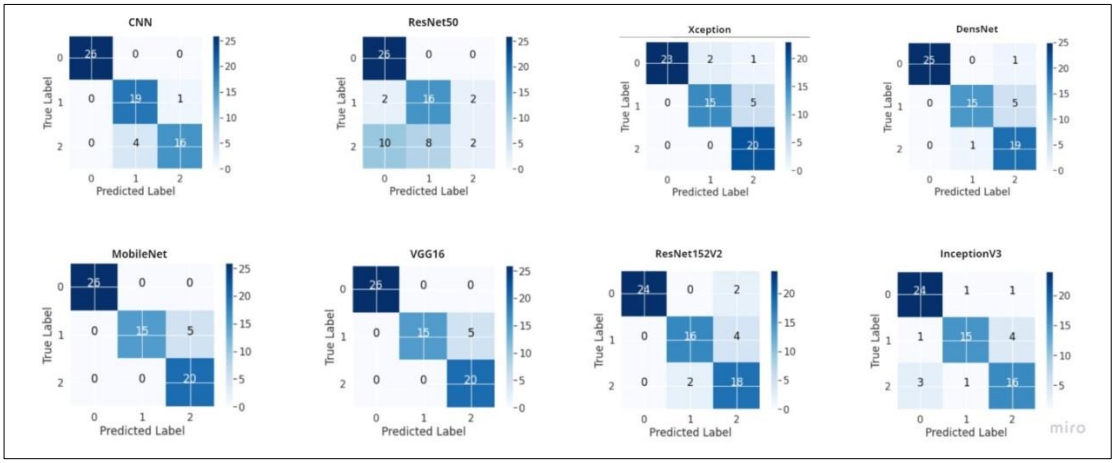


Figure 5. Confusion matrix of dl models

Precision refers to the percentage of samples from a given class that were correctly predicted out of all samples predicted as belonging to that class. Recall denotes the percentage of samples in a given class that were correctly predicted out of all actual samples from that class. The F1-score is the harmonic mean of precision and recall, providing a balanced measure of both metrics.

Table 2. Results obtained with performance metrics

Model	Class	Precision	Recall	F1-score	Specificity
CNN	Covid	1.00	1.00	1.00	1.00
	Normal	0.83	0.95	0.88	0.91
	Pneumonia	0.94	0.80	0.86	0.98
Resnet50	Covid	0.68	1.00	0.81	0.70
	Normal	0.67	0.80	0.73	0.83
	Pneumonia	0.50	0.10	0.17	0.96
Xception	Covid	1.00	0.88	0.94	1.00
	Normal	0.88	0.75	0.81	0.96
	Pneumonia	0.77	1.00	0.87	0.87
DenseNet	Covid	1.00	0.96	0.98	1.00
	Normal	0.94	0.75	0.83	0.98
	Pneumonia	0.76	0.95	0.84	0.87
MobileNet	Covid	1.00	1.00	1.00	1.00
	Normal	1.00	0.75	0.86	1.00
	Pneumonia	0.80	1.00	0.89	0.89
VGG16	Covid	1.00	1.00	1.00	1.00
	Normal	1.00	0.75	0.86	1.00
	Pneumonia	0.80	1.00	0.89	0.89
Resnet152v2	Covid	1.00	0.92	0.96	1.00
	Normal	0.89	0.80	0.84	0.96
	Pneumonia	0.75	0.90	0.82	0.87
Inceptionv3	Covid	0.86	0.92	0.89	0.90
	Normal	0.88	0.75	0.81	0.96
	Pneumonia	0.76	0.80	0.78	0.89

Class-based FPR and FNR analysis results are given in Table 3. Accordingly, MobileNet and VGG16 models provide quite reliable results, especially in Covid and Pneumonia classes. However, it was observed that samples belonging to the Normal class were incorrectly classified by up to 25% in both models. Apart from this, the ResNet50 model has high FPR and serious error rates, especially in the Pneumonia class, with 90% FNR. Therefore, it should not be considered a reliable model. In general, the DenseNet model stands out as one of the models that offers the most balanced performance in terms of both FPR and FNR. According to the statements above, the top models are CNN, MobileNet, and VGG16, which have a 92% total accuracy rate.

Table 3. Class-based fpr and fnr values for all models

Model	Class	FPR	FNR
CNN	Covid	0.0244	0.0000
	Normal	0.0870	0.0952
	Pneumonia	0.0213	0.2000
Resnet50	Covid	0.3000	0.0000
	Normal	0.1739	0.2000
	Pneumonia	0.0435	0.9000
Xception	Covid	0.0000	0.1154
	Normal	0.0435	0.2500
	Pneumonia	0.1304	0.0000
DenseNet	Covid	0.0000	0.0385
	Normal	0.0217	0.2500
	Pneumonia	0.1304	0.0500
MobileNet	Covid	0.0000	0.0000
	Normal	0.0000	0.2500
	Pneumonia	0.1087	0.0000
VGG16	Covid	0.0000	0.0000
	Normal	0.0000	0.2500
	Pneumonia	0.1087	0.0000
Resnet152v2	Covid	0.0000	0.0769
	Normal	0.0435	0.2000
	Pneumonia	0.1304	0.1000
Inceptionv3	Covid	0.1000	0.0769
	Normal	0.0435	0.2500
	Pneumonia	0.1087	0.2000

The Covid class, which is the most important class here, has the highest sensitivity and specificity for these models. On the other hand, ResNet50 is the lowest model with an accuracy 67%, the lowest sensitivity and specificity for the Pneumonia class; therefore, it is the worst model.

It is important to keep in mind that these reports barely scratch the surface in terms of how well these models did on our particular dataset; they might not do as well on other datasets or tasks. We can judge the models based on AUC-ROC or AUC-PR and the outcome of the models' performance can be optimized through the ensembling and hyperparameter tuning technique as well.

5. DISCUSSION

To evaluate the effectiveness of our proposed DL framework for classifying chest X-ray images into COVID-19, pneumonia, and normal categories, we conducted a comprehensive comparative analysis with several state-of-the-art methods from recent literature. Our results demonstrate that the proposed approach, particularly when utilizing fine-tuned versions of MobileNet, CNN, and VGG16 architectures, consistently outperforms many existing models, especially in terms of classification accuracy, generalizability, and practical applicability.

Notably, our best-performing models achieved an accuracy of 92.4% on a multi-class dataset, surpassing the performance of various studies that employed either binary classification or more computationally complex ensemble strategies. For instance, in the study by Aljawarneh et al. [42], the enhanced CNN model reached 92.4% accuracy on pneumonia detection, while ResNet50 performed considerably lower at 82.8%. Similarly, Shah et al. [31] introduced a CNN-based CT model that achieved only 82.1% accuracy for two-class classification, highlighting the limitations of CT imaging alone and the effectiveness of our CXR-based approach. Other frameworks such as the CNN-LSTM with moth flame optimization by Hamza et al. [24] reported variable performance across datasets, with the lowest accuracy at 93%, but required heavy fusion techniques that increased computational complexity and training time. In contrast, our method remains lightweight and easily deployable in clinical settings. Furthermore, while Reshan et al. [27] evaluated eight pre-trained models and found MobileNet to perform best with 94.23% and 93.75% on two datasets, their study was constrained by dataset imbalance and lacked a unified preprocessing strategy.

Our frameworks addresses these issues by implementing data normalization, class balancing, and consistent evaluation metrics. Additionally, our method maintains high precision and recall across all classes, which is crucial for minimizing false negatives in COVID-19 detection. Taken together, the comparative results clearly establish the superiority of our proposed approach in terms of classification performance, architectural efficiency, and its potential utility as a reliable diagnostic support tool in real-world medical applications.

6. CONCLUSION AND FUTURE WORKS

The research explored the performance of various DL models, including CNN, Resnet50, Xception, DenseNet, Mobilenet, VGG16, Resnet152v2, and Inceptionv3, on the classification of chest X-ray images as Covid, Normal, and Pneumonia classes for a COVID-19 recognition scenario. The COVID-19 class achieved the highest sensitivity across all models, with CNN, MobileNet, and VGG16 reaching the highest overall accuracy of approximately 92%. However, the classification performance for the Pneumonia class was notably lower, especially in the case of ResNet50, which yielded the poorest results among all models. In contrast, MobileNet and VGG16 demonstrated balanced and reliable performance across all three classes. The results indicate that computer-aided diagnostic technologies may be beneficial in the COVID-19 detection and treatment process. The results are specific to the dataset. Therefore, it is necessary to conduct more experiments to establish how well the models work for other datasets and challenges. The performance of the models can be enhanced by adding more optimization strategies, and preprocessing stages such as hyperparameters tuning, data augmentation and architecture compositions. In addition to X-ray images, CT scan images can also be studied, which can provide detailed anatomical information at higher resolution and increase accuracy.

A future path to increase the DL models' performance is by using ensemble voting for COVID-19 diagnosis through chest X-ray images. Ensemble learning is a method that uses multiple models to maximize accuracy and minimize overfitting. One kind of ensemble learning is ensemble voting, which is where the predictions of multiple models are consolidated to select the class that has the maximum votes as the final estimate.

We could apply DL models such as CNN, Resnet50, Xception, DenseNet, Mobilenet, VGG16, Resnet152v2, and Inceptionv3 and combine their predictions with a voting mechanism. Different voting processes could be attempted, such as majority voting or weighted voting, where the model which performs better or bests correctly for one class is weighted more.

Although this study focused solely on image-based model evaluation, the integration of patient clinical data and the use of hybrid approaches combining multiple models are expected to improve diagnostic accuracy. Therefore, such clinically supported and multi-model frameworks are considered promising directions for future research.

A potential avenue for future research is the exploration of transfer learning techniques. This approach entails pre-training models on a large-scale dataset before refining them for a particular objective. The core idea is to utilize insights gained from large datasets to improve performance on smaller ones, thereby reducing both

computational cost and processing time. Consequently, pre-trained models can be adapted for COVID-19 detection by fine-tuning them with a combination of a small COVID-19 dataset and a larger chest X-ray dataset, such as the ChestX-ray14.

Finally, investigation of the possible solution is to increase the precision of the diagnostic. The latter may lead to explore possible merger of imaging techniques and clinical data, for example, involving patient demographics, medical history, and laboratory findings. Thus, the information about the patient can be more beneficial and may help to examine accurately COVID.

DECLARATION OF ETHICAL STANDARDS

The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

AUTHORS' CONTRIBUTIONS

Maral A. Mustafa: Performed the experiments, analyzed and evaluated the results, and drafted the manuscript.

O. Ayhan ERDEM: Observed the results obtained, and reviewed the manuscript.

Esra SÖĞÜT: Reviewed the experiments, analyzed the results, and reviewed the manuscript.

CONFLICT OF INTEREST

There is no conflict of interest in this study.

REFERENCES

[1] Zhu, N., Zhang, D. Wang, W. Li, X. Yang, B. Song, J. Zhao, X. Huang, B. Shi, W. Lu, R., et al. "A novel coronavirus from patients with pneumonia in China, 2019", ***New England journal of medicine,*** 382(8): 727–733, (2020).

[2] Barth, R. F., Buja, L., Barth, A. L., Carpenter, D. E., and Parwani, A. V., "A Comparison of the Clinical, Viral, Pathologic, and Immunologic Features of Severe Acute Respiratory Syndrome (SARS), Middle East Respiratory Syndrome (MERS), and Coronavirus 2019 (COVID-19) Diseases", Archives of ***Pathology & Laboratory Medicine,*** 145(10): 1194–1211, (2021).

[3] Cui, J., Li, F., and Shi, Z. L., "Origin and evolution of pathogenic coronaviruses", ***Nature reviews microbiology***, 17(3): 181-192, (2019).

[4] Wong, C. K., Lau, K. T., Au, I. C., Xiong, X., Lau, E. H., and Cowling, B. J., "Clinical improvement, outcomes, antiviral activity, and costs associated with early treatment with remdesivir for patients with coronavirus disease 2019 (COVID-19)", ***Clinical Infectious Diseases***, 74(8): 1450-1458, (2022).

[5] Ravi, V., Narasimhan, H., Chakraborty, C., and Pham, T. D., "Deep learning-based meta-classifier approach for COVID-19 classification using CT scan and chest X-ray images", ***Multimedia systems***, 28(4): 1401-1415, (2022).

[6] Verma, A., Amin, S. B., Naeem, M., & Saha, M., "Detecting COVID-19 from chest computed tomography scans using AI-driven android application", ***Computers in biology and medicine,*** 143: 105298, (2022).

[7] Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., et al., "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19)", ***European radiology***, 31: 6096-6104, (2021).

[8] Nanda, A., Barik, R. C., and Bakshi, S., "SSO-RBNN driven brain tumor classification with Saliency-K-means segmentation technique", ***Biomedical Signal Processing and Control***, 81: 104356, (2023).

[9] Littrup, P. J., Freeman-Gibb, L., Andea, A., White, M., Amerikia, K. C., Bouwman, D., Harb, T., and Sakr, W., "Cryotherapy for breast fibroadenomas", ***Radiology***, 234(1): 63-72, (2005).

[10] Darıcı, M. B., "Performance analysis of combination of cnn-based models with adaboost algorithm to diagnose covid-19 disease", ***Journal of Polytechnic***, 26(1): 179-190, (2023).

[11] Zieleskiewicz, L., Markarian, T., Lopez, A., Taguet, C., Mohammedi, N., Boucekine, M., Baumstarck, K., Besch, G., Mathon, G., Duclos, G., et al., "Comparative study of lung ultrasound and chest computed tomography scan in the assessment of severity of confirmed COVID-19 pneumonia", ***Intensive care medicine***, 46: 1707-1713, (2020).

[12] Khan, E., Rehman, M. Z. U., Ahmed, F., Alfouzan, F. A., Alzahrani, N. M., and Ahmad, J., "Chest X-ray classification for the detection of COVID-19 using deep learning techniques", ***Sensors***, 22(3): 1211, (2022).

[13] Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N. C., Sardanelli, F., "AI applications to medical images: From machine learning to deep learning", ***Physica medica***, 83: 9-24, (2021).

[14] Çelikdemir, M. Y., and Akbal, A., "A Deep Learning Based on Automatic Cerebral Aneurysm Detection in Brain Computed Tomography Angiography Scan Images", ***Journal of Polytechnic***, 28(1): 147-157, (2025).

[15] Marentakis, P., Karaiskos, P., Kouloulias, V., Kelekis, N., Argentos, S., Oikonomopoulos, N., and Loukas, C., "Lung cancer histology classification from CT images based on radiomics and deep learning models", ***Medical & biological engineering & computing,*** 59: 215-226, (2021).

[16] Roy, S., Meena, T., and Lim, S. J., "Demystifying supervised learning in healthcare 4.0: A new reality of transforming diagnostic medicine", ***Diagnostics***, 12(10): 2549, (2022).

[17] Dündar Ö., and Koçer S., "Pneumonia detection from pediatric lung X-ray images using artificial neural networks", ***Journal of Polytechnic***, 27(5): 1843-1852, (2024).

[18] Kaya, Y., Yiner, Z., Kaya, M., and Kuncan, F., "A new approach to COVID-19 detection from X-ray images using angle transformation with GoogleNet and LSTM", ***Measurement Science and Technology***, 33(12): 124011, (2022).

[19] Ismael, A. M., and Şengür, A., "Deep learning approaches for COVID-19 detection based on chest X-ray images", ***Expert Systems with Applications***, 164: 114054, (2021).

[20] Yılmaz, A. "Diagnosing COVID-19 from X-Ray images with using multi-channel CNN architecture", ***Journal of the Faculty of Engineering and Architecture of Gazi Universit***y, 36(4): 1761-1774, (2021).

[21] Nair, R., Alhudhaif, A., Koundal, D., Doewes, R. I., and Sharma, P., "Deep learning-based COVID-19 detection system using pulmonary CT scans", ***Turkish Journal of Electrical Engineering and Computer Sciences***, 29(8): 2716-2727, (2021).

[22] Khan, M. A., Azhar, M., Ibrar, K., Alqahtani, A., Alsubai, S., Binbusayyis, A., Kim, Y.J., and Chang, B., "COVID-19 classification from chest X-ray images: a framework of deep explainable artificial intelligence", ***Computational Intelligence and Neuroscience***, 2022(1): 4254631, (2022).

[23] Iqbal, A., Usman, M., and Ahmed, Z., "Tuberculosis chest X-ray detection using CNN-based hybrid segmentation and classification approach," ***Biomedical Signal Processing and Control***, 84: 104667, (2023).

[24] Hamza, A., Attique Khan, M., Wang, S. H., Alqahtani, A., Alsubai, S., Binbusayyis, A., Hussein, H.S., Martinetz, T.M., and Alshazly, H., "COVID-19 classification using chest X-ray images: A framework of CNN-LSTM and improved max value moth flame optimization", ***Frontiers in Public Health***, 10: 948205, (2022).

[25] Fan, X., Feng, X., Dong, Y., and Hou, H., "COVID-19 CT image recognition algorithm based on transformer and CNN", ***Displays***, 72: 102150, (2022).

[26] Garg, A., Salehi, S., La Rocca, M., Garner, R., and Duncan, D., "Efficient and visualizable convolutional neural networks for COVID-19 classification using Chest CT", ***Expert Systems with Applications***, 195: 116540, (2022).

[27] Reshan, M. S. A., Gill, K. S., Anand, V., Gupta, S., Alshahrani, H., Sulaiman, A., and Shaikh, A., "Detection of pneumonia from chest X-ray images utilizing mobilenet model", ***Healthcare***, 11(11): 1561, (2023).

[28] AbdElhamid, A. A., AbdElhalim, E., Mohamed, M. A., and Khalifa, F., "Multi-classification of chest X-rays for COVID-19 diagnosis using deep learning algorithms", ***Applied Sciences***, 12(4): 2080, (2022).

[29] Gampala, V., Rathan, K., S, C. N., Shajin, F. H., and Rajesh, P., "Diagnosis of COVID-19 patients by adapting hyper parametertuned deep belief network using hosted cuckoo optimization algorithm", ***Electromagnetic Biology and Medicine***, 41(3): 257-271, (2022).

[30] Thakur, S., and Kumar, A., "X-ray and CT-scan-based automated detection and classification of covid-19 using convolutional neural networks (CNN)", ***Biomedical Signal Processing and Control***, 69: 102920, (2021).

[31] Shah, V., Keniya, R., Shridharani, A., Punjabi, M., Shah, J., and Mehendale, N., "Diagnosis of COVID-19 using CT scan images and deep learning techniques", ***Emergency Radiology***, 28: 497-505, (2021).

[32] Muhammad, G., and Hossain, M. S., "COVID-19 and non-COVID-19 classification using multi-layers fusion from lung ultrasound images", ***Information Fusion***, 72: 80-88, (2021).

[33] Hussain, E., Hasan, M., Rahman, M. A., Lee, I., Tamanna, T., and Parvez, M. Z., "CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images", ***Chaos, Solitons & Fractals***, 142: 110495, (2021).

[34] Ezzat, D., Hassanien, A. E., and Ella, H. A., "An optimized deep learning architecture for the diagnosis of COVID-19 disease based on gravitational search optimization", ***Applied Soft Computing***, 98: 106742, (2021).

[35] Nandhini Abirami, R., Durai Raj Vincent, P. M., Rajinikanth, V., and Kadry, S., "COVID-19 classification using medical image synthesis by generative adversarial networks", ***International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems***, 30(03): 385-401, (2022).

[36] Abirami, N., Vincent, D. R., and Kadry, S., "P2P-COVID-GAN: Classification and segmentation of COVID-19 lung infections from CT images using GAN", ***International Journal of Data Warehousing and Mining***, 17(4): 101-118, (2021).

[37] Khan, I. U., and Aslam, N., "A deep-learning-based framework for automated diagnosis of COVID-19 using X-ray images", ***Information***, 11(9): 419, (2020).

[38] Ullah, N., Khan, J. A., Almakdi, S., Khan, M. S., Alshehri, M., Alboaneen, D., and Raza, A., "A novel CovidDetNet deep learning model for effective COVID-19 infection detection using chest radiograph images", ***Applied Sciences***, 12(12): 6269, (2022).

[39] Arman, S. E., Rahman, S., and Deowan, S. A., "COVIDXception-Net: A Bayesian optimization-based deep learning approach to diagnose COVID-19 from X-Ray images", ***SN Computer Science***, 3(2): 115, (2022).

[40] Abad, M., Casas-Roma, J., and Prados, F., "Generalizable disease detection using model ensemble on chest X-ray images", ***Scientific Reports***, 14(1): 5890, (2024).

[41] Singh, T., Mishra, S., Kalra, R., Satakshi, Kumar, M., and Kim, T., "COVID-19 severity detection using chest X-ray segmentation and deep learning", ***Scientific Reports***, 14(1): 19846, (2024).

[42] Aljawarneh, S. A., and Al-Quraan, R., "Pneumonia detection using enhanced convolutional neural network model on chest x-ray images", ***Big Data***, 13(1): 16-29, (2025).

[43] Prince, R., Niu, Z., Khan, Z. Y., Chambua, J., Yousif, A., Patrick, N., and Jennifer, B., "Interpretable COVID-19 chest X-ray detection based on handcrafted feature analysis and sequential neural network***", Computers in Biology and Medicine***, 186: 109659, (2025).

[44] El-Ghandour, M., and Obayya, M. I., "Pneumonia detection in chest x-ray images using an optimized ensemble with XGBoost classifier", ***Multimedia Tools and Applications***, 84(9): 5491-5521, (2025).

[45] ***https://www.sirm.org/category/senza-categoria/covid-19/***, "COVID-19 Database", (2020).